1. Lenguajes formales

Juan B. Castellano
Desp 2210
jcastellanos@fi.upm.es

Definiciones

Alfabeto

Un conjunto finito de símbolos.

Por ejemplo:

- $\sum_1$={0,1}
- $\sum_2$={a,b,c,d}
- $\sum_3$={triangulo, cuadrado, punto}

Lenguaje universal

Dado un alfabeto ∑, se define lenguaje universal como

∑*={todas las palabras que se pueden formar con elementos de ∑, incluyendo λ}

$\lambda \in \sum^*$ para cualquier ∑

Subalfabeto

Es un conjunto de un alfabeto

Por ejemplo:

- $\sum_4$={a,c} subalfabeto de $\sum_2$

Palabra

Dado un alfabeto ∑, una palabra x, es una sucesión ordenada de símbolos de ∑, con o sin repetición.

Por ejemplo:

- En $\sum_2$={a,b,c,d}, x e y son palabras sobre $\sum_2$ → x=bbad, y = adc

Los elementos de ∑ los designamos con a,b,c,d… y las palabras las representamos por x,y,z,t…

Longitud de una palabra

Dado ∑, una palabra de x tiene longitud n, si n es el número de signos de la palabra.

Por ejemplo:

- X=abcda $|x|=5$
- Y=aaa $|y|=3$

Palabra λ

Esta palabra no tiene símbolos y no pertenece a ningún alfabeto.

$|\boldsymbol{\lambda}|=0$

Los símbolos del alfabeto, que también son palabras, tienen longitud 1.

Por ejemplo:

$\sum=\{0,1\} \rightarrow$ 0 y 1 son símbolos y palabras. $|0|=1$ y $|1|=1$

Palabra inversa

Dado un alfabeto ∑ y x es una palabra sobre ∑, la palabra inversa de x, x^{-1} es la palabra que tiene los símbolos de x puestos en sentido inverso. Sin embargo esta no es la definición más formal, esta sí:

Dada x palabra sobre ∑, definimos la palabra inversa según el tamaño de las palabras

1. Si $|x|=0$ si $x=\lambda \rightarrow x^{-1}=\lambda$
2. Si $|x|>0$, Existe w palabra y Existe a símbolo tal que $x=wa \rightarrow x^{-1}=aw^{-1}$

Por ejemplo:

$\sum=\{c,d,e\} \rightarrow x=cdec$,, $x^{-1}=cedc$

Palabra simétrica

Dado ∑ y una palabra $x\epsilon\sum^*$(ver siguiente), x es simétrica si $x=x^{-1}$

Por ejemplo:

$\sum=\{0,1\}$, x= 0110 $x=x^{-1}$

Operaciones con palabras

Concatenación de palabras

$\cdot : \sum^* x \sum^* \rightarrow \sum^*$
X, y $\rightarrow x\cdot y$

Si x=ab, y=cd $\rightarrow x\cdot y=abcd$

Propiedades

1. Operación bien definida
   a. Para todo x,y ϵ ∑* x·y es una única palabra ϵ ∑*
2. Es asociativa
   a. Para todo x,y,z ϵ ∑* (x·y)·z=x·(y·z)
3. No es conmutativa
   a. Para todo x,y ϵ ∑* x·y≠y·x
4. Elemento neutro
   a. Para todo x ϵ ∑* Existe neutro λ tal que x·λ=λ·x=x

(∑*, ·) Tiene estructura de semigrupo con elemento neutro

Propiedades simplificativas

1. X,y,z ϵ ∑* x·y=x·z => y=z x·y=z·y => x=z
2. Dado ∑, dados x,y ϵ ∑* se cumple que $(x\cdot y)^{-1}=y^{-1}\cdot x^{-1}$
   a. Demostración mediante inducción en el tamaño y
      i. Hay que demostrar que se cumple la propiedad para la palabra más pequeña

         $(x\cdot y)^{-1}=y^{-1}\cdot x^{-1}$ |y|=0 => y=λ

         $(x\cdot y)^{-1}=(x\cdot\lambda)^{-1}=x^{-1}$

         $y^{-1}\cdot x^{-1}=\lambda^{-1}\cdot x^{-1}=\lambda\cdot x^{-1}=x^{-1}$

         Por tanto se cumple.
      ii. Suponemos que se cumple la propiedad para todas las palabras y con tamaño n, es decir, para todo x ϵ ∑* y para todo y ϵ ∑* con |y|=n se cumple $(x\cdot y)^{-1}=y^{-1}\cdot x^{-1}$
      iii. Hay que demostrar que para todo z con tamaño n+1 se cumple la propiedad, es decir, que para todo x ϵ ∑* y para todo z ϵ ∑* con |z|=n+1 se cumple $(x\cdot z)^{-1}=z^{-1}\cdot x^{-1}$

         Si |z|=n+1, existe y ϵ ∑* |y|=n y existe a ϵ ∑ tal que z=ya

         $(xz)^{-1}=(x\cdot(ya))^{-1}=((xy)\cdot a)^{-1}=a\cdot(xy)^{-1}=a\cdot y^{-1}\cdot x^{-1}=(ay^{-1})\cdot x^{-1}=\boxed{z^{-1}\cdot x^{-1}}$

         asociativa

         Def de palabra inversa

         Hipótesis de inducción

         asociativa

         Def. z^{-1}

Potencia

∑, alfabeto

x∈∑*

potencia i-ésima $x^i=x\cdot x\cdot x\cdots x$, ej: $x^3=xxx$

Def.- $x^0=\lambda$

$x^1=x$

$|x^1|=i|x|$

$\sum^+=\sum^*-\{\lambda\}$

$\sum^+$ se puede obtener por concatenación de potencias

Lenguaje

Dado un alfabeto $\sum$, un lenguaje L es un conjunto de palabras sobre $\sum$ que cumplen una propiedad:

$L = \{x\epsilon\sum^* | x \text{ cumple } P\}$

Por ejemplo:

- Si $\sum=\{0,1\}$

  $L = \{0,01,011,0111,\ldots\} = \{x\epsilon\{0,1\}^* |$ son palabras cuyo 1er símbolo es 0 y el resto 1's, si los tiene$\}$

- Si $\sum$, alfabeto cualquiera

  $L_1=\{x\epsilon\sum^* |\ |x|=i\}$

  $L_2= \{ x\epsilon\sum^* |\ |x|=i+1\}$

Lenguaje $\mathcal{L}$

$\mathcal{L}$={Lenguajes sobre $\sum$}

Lenguaje inverso

Dado L sobre $\sum$,

$L^{-1}=\{x\epsilon\sum^* | \text{existe } z \epsilon L \text{ y } x=z^{-1}\}$

Lenguaje complementario de L

$L^{-}=\{x\epsilon\sum^* | x \text{ no}\epsilon\ L\}$

Operaciones con lenguajes

Unión de lenguajes

$\sum$, alfabeto,, L_1, L_2 lenguajes. L_1, L_2 contenido en $\sum^*$

U: $\sum^* x \sum^* \rightarrow \sum^*$
$L_1, L_2 \rightarrow L_1 U L_2$

$L_1 U L_2=\{x\epsilon \sum^* | x\epsilon L_1 \text{ ó } x\epsilon L_2\}$

Propiedades

1. Está bien definida
   a. L_1, L_2 ϵ $\mathcal{L}$ L_1unionL_2 contenido en $\sum^*$ es otro lenguaje.
2. Es asociativa
   a. Para todo L_1, L_2, L_3 $L_1U(L_2UL_3)= (L1UL_2)UL3$
3. Es conmutativa

   a. Para todo L_1,L_2 $L_1 U L_2 = L_2 U L_1$
4. Elemento neutro, Φ
   a. Para todo L∊$\mathcal{L}$, existe neutro tal que LUΦ=ΦUL=L

($\mathcal{L}$, U) -> Tiene estructura de semigrupo conmutativo elemento neutro.

Concatenación

∑, $\mathcal{L}$={lenguajes sobre ∑)

$\cdot: \mathcal{L} \times \mathcal{L} \rightarrow \mathcal{L}$
$L_1, L_2 \rightarrow L_1 \cdot L_2$

$L_1 \cdot L_2$={x|x=z·y donde z∊L_1 e y∊L_2}

Propiedades

1. Ley de composición interna
2. Asociativa
3. No conmutativa
4. Elemento neutro
   i. {λ} no confundir con λ, uno es lenguaje el otro palabra.

($\mathcal{L}$,·) Semigrupo con elemento neutro

Potencia de lenguaje

L sobre ∑,

L^i= L·L·L···L, i veces

Def. de cierre de un lenguaje. Estrella de Kleene

Dado L,

$L^*=L^0 U L^1 U L^2 U L^3 U \cdots$

L^*= U(de 0 a infinito) L^i

La clausura positiva es L^+ (L^+= L^*-{λ})

L^+= U(de 1 a infinito) L^i

Expresiones y Lenguajes regulares

Expresiones regulares

B = ∑U{Φ, λ, ·, *, ()}

Expresiones regulares sobre $\sum$ α, $\alpha \epsilon$Exp Reg($\sum$) son las palabras que se pueden formar con B de manera recursiva:

i) Φ, >, a$\epsilon\sum$ son ER($\sum$)

ii) α, $\beta\epsilon$ ER($\sum$) => $\alpha+\beta\epsilon$ER($\sum$)

iii) α, $\beta\epsilon$ ER($\sum$) => $\alpha\cdot\beta\epsilon$ER($\sum$)

iv) $\alpha\epsilon$ ER($\sum$) => $\alpha^*\epsilon$ER($\sum$)

v) $\alpha\epsilon$ ER($\sum$) => (α)ϵER($\sum$)

vi) Son Expresiones regulares sobre $\sum$ todas las concatenaciones de las ER anteriores.

Por ejemplo:

1.- $\sum=\{0,1\}$ L={x$\epsilon\sum^*$/en x aparece el 1 dos o tres veces, la 1ª y 2ª de las cuales no son consecutivos}

α= 0* 1 0 0* 1 0*(1 0* + λ)

Explicación: un 0 con asterisco * significa que aparece un 0 en todas las veces que quieras, inclusive ninguna. El paréntesis indica que lo que hay dentro puede o no estar.

2.- $\sum=\{a,b\}$ $L_2=\{x\epsilon\sum^*/|x|_a$ es impar}

α=(b* a b* a b*)* b* a b* aϵL_2 ba ϵL_2 bab^3 ϵL_2

3.- $\sum=\{a,b\}$ $L_3=\{a^nb^m/ n, m>0\}$

α= aa*bb*

4.- $\sum=\{a, b\}$ $L_6=\{a^nb^n/n>0\}$

α=a(ab)*b no pertenece a L_6, aunque lo parezca en este caso α no sería una expresión regular, ya que no mantiene el orden. Esto es porque L_6 no es un lenguaje regular.

Lenguajes regulares

Lenguaje regular sobre $\sum$, es una aplicación L:

L:ER($\sum$) → $\delta\sum$=Parte de ($\sum^*$)
α → L(α) de manera

a) λ, Φ, $\mu \in \sum \in$ ER($\sum$) => L(λ)={λ} L(Φ)=Φ L(μ)={μ}

b) α, $\beta \in$ ER($\sum$) => L($\alpha+\beta$)=L(α)UL(β)

c) α, $\beta \in$ ER($\sum$) => L($\alpha\cdot\beta$)=L(α)·L(β)

d) $\alpha \in$ ER($\sum$) => L(α^*)=(L(α^*))*

Lema.- L sobre $\sum$ es regular <-> Existe $\alpha \epsilon ER(\sum)/L(\alpha)=L$

Ejercicio

Con $L=\{0^r1^s/r,s>=1\}$

Encontrar el complementario de L como unión de tres lenguajes disjuntos y dar una expresión regular de ella.

L_1={palabras que solo tienen 0's}={0}*

L_2={palabras que empiezan por 1}= $1\sum^*$

$L_3=\{0^x1^s0\ \sum^*\}$

$L(\alpha)=L1$ $L(\beta)=L2$ $L(\mu)=L3$

$\alpha=0^*$ $\beta=1(0+1)^*\ \sum^*$ $\mu=00^*11^*0(0+1)^*$

$\alpha+\beta+\mu=0^*+1(0+1)^*+00^*11^*0(0+1)^*$

Prefijo o cabeza

u $\epsilon\sum^*$ es prefijo o cabeza de $v\epsilon\sum^*$ si

Existe x $\epsilon\sum^*$ ux=v

Sufijo o cola

u $\epsilon\sum^*$ es sufijo o cola de v si

Existe x $\epsilon\sum^*$ xu=v

Propiedades

$\alpha, \beta\epsilon\ ER(\sum)$ α y β son equivalentes = $L(\alpha) = L(\beta)$

-> + es conmutativa y asociativa

$\alpha+\beta=\beta+\alpha$ $\alpha, \beta \in ER(\sum)$

$(\alpha+\beta)+\mu=\alpha+(\beta+\mu)$ $\alpha, \beta, \mu \in ER(\sum)$

-> · es asociativa y distributiva respecto de +

$(\alpha\cdot\beta)\cdot\mu=\alpha\cdot(\beta\cdot\mu)$ $\alpha, \beta, \mu \in ER(\sum)$

$\alpha\cdot(\beta+\mu)=\alpha\cdot\beta+\alpha\cdot\mu$ $\alpha, \beta, \mu \in ER(\sum)$

$(\beta+\mu)\cdot\alpha=\beta\cdot\alpha+\mu\cdot\alpha$ $\alpha, \beta, \mu \in ER(\sum)$

-> Φ es neutro respecto de +

$\alpha+\Phi=\Phi+\alpha=\alpha$ $\alpha \in ER(\sum)$

-> λ es neutro respecto de ·

$\lambda\cdot\alpha=\alpha\cdot\lambda=\alpha$ $\alpha \in ER(\sum)$

-> $\Phi\cdot\alpha=\alpha\cdot\Phi=\Phi$ $\alpha \in ER(\sum)$

-> $\Phi^*=\lambda$

-> $\lambda^*=\lambda$

-> $\alpha^*=\lambda+\alpha+\alpha^2+\alpha^3+\cdots+\alpha^u$ $\alpha \in ER(\Sigma)$

-> $\alpha^*=\lambda+\alpha+\alpha^2+\alpha^3+\cdots+\alpha^u+\alpha^{u+1}+\alpha^*$ $\alpha \in ER(\Sigma)$

-> $\alpha^*\cdot\alpha^*=\alpha^*$ $\alpha \in ER(\Sigma)$

-> $(\alpha^*)^*=\alpha^*$ $\alpha \in ER(\Sigma)$

-> $\boldsymbol{\alpha=\lambda+\alpha\cdot\alpha^*}$ $\alpha \in ER(\Sigma)$

-> $\alpha\cdot\alpha^*=\alpha^*\cdot\alpha$ $\alpha \in ER(\Sigma)$

-> $(\alpha^*+\beta^*)^*=(\alpha+\beta)^*=(\alpha^*\cdot\beta^*)^*$ $\alpha, \beta \in ER(\Sigma)$

-> $(\alpha\cdot\beta)^*\cdot\alpha=\alpha\cdot(\beta\cdot\alpha)^*$ $\alpha, \beta \in ER(\Sigma)$

-> $(\alpha^*\cdot\beta)^*\cdot\alpha^*=(\alpha+\beta)^*$ $\alpha, \beta \in ER(\Sigma)$

-> L regular → L^{-1} regular

Ejercicios

Ejer. 1.- Probar que las dos expresiones regulares son equivalentes

$E_1=a\cdot a^*\cdot(a^*\cdot b)^*+a\cdot a^*(a^*+b^*)+\lambda$

$E_2=a^*\cdot(a^*\cdot b^*)+a^*\cdot a\cdot(a+b)^*\cdot b+\lambda$

Ejer. 2.- Construir directamente una ER que represente estos lenguajes:

$L_1=\{x\in\{a,b,c\}^*/ac$ no es parte de $x\}$

$L_2=\{x\in\{a,b,c\}^*/x$ tiene un número par de ocurrencias de $ac\}$

Gramáticas formales

Consideraciones previas

$\Sigma \rightarrow \Sigma^*$

Producciones, Leyes de escritura sobre ∑

(u, v)= u::=v pares ordenados / $u, v \in \Sigma^*$

Derivación directa

x u y → x v y es derivación directa siendo(u ::=v)∊ P

z → t (z deriva directa a t)

$z, t \in \Sigma^*$ x u y → x v y $x, y \in \Sigma^*$ $u, v \in \Sigma^*$ (u ::= v)∊P

Evidentemente u u::=v → u -> v pues x,y=λ

Derivación

u $\rightarrow_{+}$v= u deriva a v o que v deriva de u = Existe $u_0, u_1, \ldots, u_m \epsilon \sum^*$

$u=u_0 \rightarrow u_1$ “ = …

“$=u_1 \rightarrow u_2$ “ $= u_{n-1} \rightarrow u_n=v$

“$=u_2 \rightarrow u_3$

$$u \rightarrow_{*} v = \begin{cases} u = v \\ o \\ u \rightarrow_{+} v \end{cases}$$

Ejemplo:

u1 = r x t u2= r y t (x::=y)ϵP

Ejemplo

$\sum$ = {A, B, C, …, x, y, z}

BA ::= ME

CA<u>BA</u>LLO → CA<u>ME</u>LLO

Gramáticas formales

$G=<\sum_T , \sum_N , S , P >$

$\sum_T$: Alfabeto de símbolos terminales

$\sum_N$: Alfabeto de símbolos no terminales (auxiliares)

S: Axioma de G ϵ $\sum_N$

P: Conjunto de producciones

u ::= v / u = xAy x,yϵ$\sum^*$ A ϵ$\sum_N$

Forma sentencial de G = { x ϵ$\sum^*$/S $\rightarrow_{*}$ x}

Sentencia de G = {x ϵ$\sum_T^*$/S $\rightarrow_{*}$ x}

Lenguaje generado por una gramática G = L(G)= Conjunto de sus sentencias

L(G)={x ϵ$\sum^*$/ S $\rightarrow_{*}$ x}

Para tener lenguaje tengo que derivar desde el axioma.

Ejemplo 1

G = <{0, 1, 2, 3, 4, …, 9}, {N,C}, N, P>

P = { N::=NC/c , c::= 0/1/2/3/…/9}

L(G) = {N}

Entonces, para crear 2015 ϵ N

N → NC → N5 → NC5→N15 → NC15 → N015 → C015 → 2015

Ejemplo 2

G=<{0,1}, {S}, S, P>

P = { S::=00S11/λ}

L(G) = {0^{2n} 1^{2n} / u>= 0}

S → 00 S 11 → 00 λ 11= $0^2 1^2$

Ejemplo 3

G=<{a,b} , {A, B, C}, A, P>

$$P=\begin{cases} A::=aABC/abC \\ CB::=BC \\ bB::=bb \\ bC::=b \end{cases}$$

L(G)=¿?

Ejemplos

1) ∑={a, b} y L∊∑* definido recursivamente así:

   i) λ∊L

   ii) a x b ϵ L y b x a ϵ L

   x∊∑*

   iii) x, y ϵ L => xy∊L

   iv) Sólo son palabras de λ lo formados por aplicación un nº finito de veces de las reglas i), ii) y iii). Definir razonadamente L.

2) ∑={a,b} L={x∊{a,b}* / $|x|_a$no=$|x|_b$}

   Solución: Sea ∑ cualquiera y L_1,L_2 contenido ∑* / L_1contenido L_2 => L_1* contenido L_2* (+)

   En nuestro caso {a, b}=L_1 => L_1*={a, b}*= ∑* contenido L* (1)

   Pero por otra parte si L contenido ∑* (L cualquiera) L* contenido ∑*

   En nuestro caso L* contenido {a,b}* (2)

   De 1 y 2 saco que L*=∑* ya que … A contenido B y B contenido A =>

No terminal

Recursividad

Una gramática G es recursiva si existe una derivación de la forma A $\rightarrow_+$ xAy donde A ϵ $\sum_n$, x,y∊∑* (pero x, y no pueden ser λ a la vez).

Si $A \rightarrow_+ Ay$ La G es recursiva izq.

Si $A \rightarrow_+ xA$ La G es recursiva der.

Una producción es recursiva si es de la forma:

$uAv ::= uxAyv$

con $A \in \sum_n$, $u,v,x,y \in \sum^*$ no todas λ

Gramáticas tipo 0

$(u ::= v) \in P$

$u = xAy$

$A \in \sum_n \quad x,y \in \sum^*$

$v \in \sum^*$(puede ser λ)

nota

Hay autores que la definen como de “estructura de frases”

** $xAy ::= xvy$

$x,y \in \sum^* \qquad A \in \sum_n$

$v \in \sum^*$(puede ser λ)

TH.- Para toda G_0 existe G_0' de estructura de frases equivalentes = $L(G_0) = L(G_0')$

Las G_0 pueden ser comprensoras => $y \rightarrow x$ puede ser $|x| \leq |y|$ la longitud de x puede ser menor (piénsese que puede ser λ)

Ejemplo 1

$$S \begin{cases} A ::= aABC/abc \\ CB ::= BC \\ bB ::= bb \\ bC ::= b \end{cases}$$

$G=<\{a,b\}, \{A,B,C\}, A, P>$

$CB ::= BC \rightarrow$ [¿Es necesaria? No]

producción de orden P

$\rightarrow$ $L(G) = \{a^u\, b^u \,/\, u \geq 1$

G no es de estructura de frases pues $CB ::= BC$ no es de la forma **

$$CB::=BC \quad \asymp \left\{ \begin{array}{l} CB ::= CX \\ CX ::= BX \\ BX ::= BC \end{array} \right\} (+)$$

entonces: $G' = <\{a,b\}, \{A,B,C\}, A, P>$

$P' = P \cup \{+\} \quad L(G)= L(G')$

Ejemplo 2

$G = < \{a,b,c\}, \{S,X\}, S, P> \qquad G=<\sum_T, \sum_N, S, P>$

$$\left\{ \begin{array}{l} S::= aSb \\ \\ \end{array} \right. \qquad \text{No de Est. de frases} \left. \right\}$$

$L(G) = \{ a^u\, c\, b^u \,/\, u \geq 0\}$

$G'=<\{a,b,c\} / \{S\}, S, P'>$

$P' = \; S::= abb/c \quad L(G) = L(G^*)$

$$P \quad \left.\begin{array}{l} Sb::= bX \\ abx::=c \end{array}\right\}$$

Gramáticas tipo 1 = $G_1 \rightarrow L_1$ = lenguajes dependientes del contexto